

## ELMo



## ELMo Tasks: Sentiment

this movie was **great!** would **watch again** **+**  
the movie was **gross** and **overwrought**, but I **liked** it **+**  
this movie was **not** really very **enjoyable** **-**

- Why are context-dependent embeddings useful here?

Peters et al. (2018)



## ELMo Tasks: NER

B-PER I-PER O O O B-LOC O O O B-ORG O O  
**Barack Obama** will travel to **Hangzhou** today for the **G20** meeting .  
PERSON LOC ORG

- Why are context-sensitive embeddings useful here?

Peters et al. (2018)



## ELMo Tasks: SRL

Gold **ARG1** **V** **ARG2** **ARG3**  
Housing starts are expected to quicken a bit from August's pace

- Identify arguments of the verb
- Why are context-dependent embeddings useful here?

Peters et al. (2018)



## ELMo

TASK	PREVIOUS SOTA		OUR BASELINE	ELMo + BASELINE	INCREASE (ABSOLUTE/ RELATIVE)
SQuAD	Liu et al. (2017)	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al. (2017)	88.6	88.0	88.7 ± 0.17	0.7 / 5.8%
SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al. (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al. (2017)	91.93 ± 0.19	90.15	92.22 ± 0.10	2.06 / 21%
SST-5	McCann et al. (2017)	53.7	51.4	54.7 ± 0.5	3.3 / 6.8%

- ▶ Training: 1B words of text taken from MT data (news, etc.)
- ▶ SST5 = sentiment, SQuAD = QA (coming later)
- ▶ Large gains across many tasks

Peters et al. (2018)



## ELMo

	Source	Nearest Neighbors
GloVe	play	playing, game, games, played, players, plays, player, Play, football, multiplayer
biLM	Chico Ruiz made a spectacular <u>play</u> on Alusik 's grounder {...}	Kieffer , the only junior in the group , was commended for his ability to hit in the clutch , as well as his all-round excellent play .
	Olivia De Havilland signed to do a Broadway <u>play</u> for Garson {...}	{...} they were actors who had been handed fat roles in a successful <u>play</u> , and had talent enough to fill the roles competently , with nice understatement .

- ▶ Nearest neighbors of words in context

Peters et al. (2018)



## ELMo

- ▶ Highly flexible: can use as an embedding layer for nearly any task
- ▶ Can be big and slow: big LSTMs can suck up GPU memory
- ▶ Strictly better than word2vec if you can afford it
- ▶ AllenNLP: calling ELMo is just one line, plenty of models for different tasks built on top of it

Peters et al. (2018)